IEICE TRANS. INF. & SYST., VOL.E96–D, NO.6 JUNE 2013

PAPER

An Accurate User Position Estimation Method Using a Single Camera for 3D Display without Glasses

Byeoung-su KIM[†a)], ***Student Member***, **Cho-il LEE**[†], **Seong-hwan JU**[††], ***and*** **Whoi-Yul KIM**[†], ***Nonmembers***

SUMMARY 3D display systems without glasses are preferred because of the inconvenience wearing of special glasses while viewing 3D content. In general, non-glass type 3D displays work by sending left and right views of the content to the corresponding eyes depending on the user position with respect to the display. Since accurate user position estimation has become a very important task for non-glass type 3D displays, most of such systems require additional hardware or suffer from low accuracy. In this paper, an accurate user position estimation method using a single camera for non-glass type 3D display is proposed. As inter-pupillary distance is utilized for the estimation, at first the face is detected and then tracked using an Active Appearance Model. The pose of face is then estimated to compensate the pose variations. To estimate the user position, a simple perspective mapping function is applied which uses the average of the inter-pupillary distance. For accuracy, personal inter-pupillary distance can also be used. Experimental results have shown that the proposed method successfully estimated the user position using a single camera. The average error for position estimation with the proposed method was small enough for viewing 3D contents.

key words: *face detection, tracking and recognition, pose estimation, active appearance models*

1. Introduction

With the rapid growth of 3D contents, 3D display systems have become more common in areas such as 3D television, cinema, and games. The most common and practical 3D display system is a stereoscopic display that requires special glasses such as polarized glasses [1], [2] or shutter glasses [3]. Since these systems are fairly simple to utilize, they have been adapted in commercial applications such as cinemas. However, wearing glasses must have been hassle to many users [4].

Preferred by most users, numerous non-glasses methods have been proposed [5]–[7]. Such technologies mostly utilize a lenticular lens or a parallax barrier. Left and right images are sent to the corresponding eyes either by blocking signals or by using cylindrical lenses as illustrated in Fig. 1. However, these systems have a limited viewing angle and limited viewing position.

To overcome this drawback, user position estimation methods have been investigated for non-glass type 3D displays using cameras [8], [9]. In these methods, different perspective views are displayed depending on the user position. In such cases, when position estimation fails, visual discomfort may occur for some people in the form of interview, crosstalk or double image [10], [11]. Therefore, accurate user position estimation becomes essential.

Many reliable and accurate methods have been proposed to estimate user position. Depth-based methods are often used to estimate user position by means of the disparity map [12], [13]. Recently, several depth sensors are developed to estimate the disparity map like Microsoft Kinect [14] and Time-Of-Flight (TOF) [15] sensors. In these methods, the accurate and direct position of a user can be estimated. However, depth-based methods are expensive and have many restrictions, such as high computational complexity or requirement of additional equipment, which may not be adequate especially for small-sized display device.

Unlike the depth-based methods, vision-based methods are more preferred as they are able to detect & estimate position and pose of user face using a single camera only [16], [17]. However, the user's position information with respect to the display must be predefined.

In order to have a solution similar as to the depth-based methods, a position estimation scheme has been proposed for vision-based methods by [18]. In this method, Information of known size or width of an object at certain distance is utilized to estimate the distance, for example, either width of face or inter-pupillary distance. However, the accuracy of the information tends to be less as compared to the depth-based methods due to variations in pose and scale.

In this paper, a proposal for an accurate user position estimation method has been made for non-glass type 3D displays using a single camera. Once a face is found, an Active Appearance Model (AAM) which is robust to various variations such as poses and face expressions is used to track the

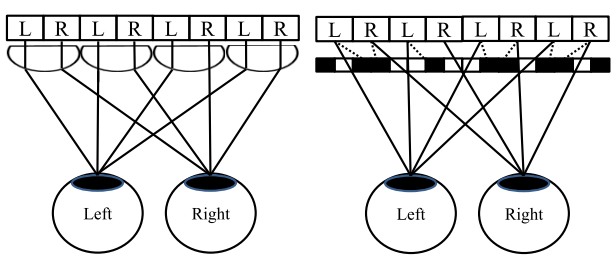


Fig. 1 An illustration of 3D display devices. (a) A lenticular lens. (b) A parallax barrier.

Manuscript received August 16, 2012.
Manuscript revised February 1, 2013.

[†]The authors are with the Electronics and Computer Engineering Department, Hanyang University, Seoul, 139–791, Republic of Korea.
[††]The author is with the LG Display, Gyeonggi-do, Republic of Korea.
a) E-mail: bskim@vision.hanyang.ac.kr

DOI: 10.1587/transinf.E96.D.1344

Copyright © 2013 The Institute of Electronics, Information and Communication Engineers

face. A face pose estimation method is then used to compensate for the error due to face rotation. In order to further improve the accuracy, user irises are detected by means of an iris detecting operator. The relative position of a user with respect to the 3D display is then easily determined by a simple relationship between the pixel numbers of inter-pupillary distance in image and distance from camera in real-world. For the position estimation, a mapping-based calibration procedure is applied using the inter-pupillary distance.

The paper is organized as follows: face tracking and pose estimation are described in Sect. 2, while position estimation is presented in Sect. 3. Experimental results are given in Sect. 4, and conclusions are presented in Sect. 5.

2. Facial Feature Tracking and Pose Estimation

In order to calculate the inter-pupillary distance, the user's face is detected and tracked first using facial features followed by the application of iris detection algorithm. Face pose is then estimated to incorporate variations in pose. Finally, user position in real-world is estimated using a simple perspective mapping function. This can be explained by Fig. 2.

2.1 Facial Feature Tracking

A face is a highly variable, deformable object, and manifests itself very differently in images depending on pose, lighting, expression, and the identity of the person. To deal with this variation, a model-based approach is particularly suited to detect and track faces in images. In the proposed method, AAM is applied for facial feature detection and is widely used for matching and tracking faces [19], [20].

In AAM algorithm, 68 facial feature points are used to align the face profile, as shown in Fig. 3 (a). The result of face tracking on a deformed face image is shown in Fig. 3 (b), which shows the robustness to various changes (like rotation, translation and barrel distortion). However, the computational complexity of AAM is very high for real time operation. For this reason, a face detection algorithm is applied to locate the initial region of AAM. A Haar-like feature-based method [21] is adopted to detect the frontal face. An integral image technique [22] is used along with this scheme to calculate the average intensity and to reduce the computational complexity.

Tracking facial features using AAM in every frame still requires a great deal of computation even if search area of the face is limited. In dealing with this problem, a feature point-tracking algorithm is utilized [23], [24]. An AAM algorithm only operates when facial feature tracking fails (movements of the face are large). To track these facial feature points, Lucas and Kanade's optical flow method is used [25].

2.2 Iris Detection

User position estimation in the proposed method makes use

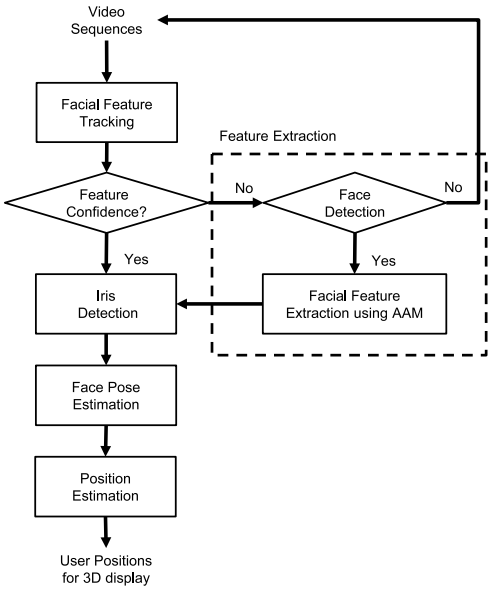
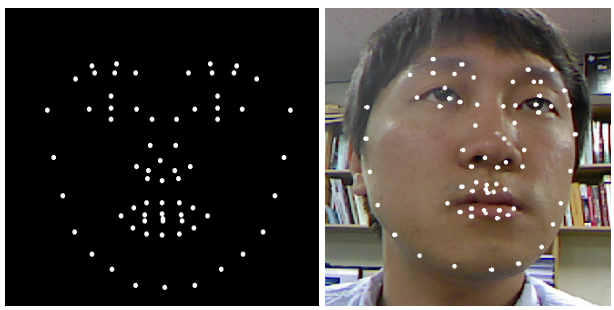


Fig. 2 The flowchart of the proposed method.



(a) (b)

Fig. 3 Example of face image labeled with facial feature points. (a) 68 facial feature points. (b) Labeled face image.

of the number of pixels between the irises called *inter-pupillary pixel distance* or inter-pupillary distance in short. Iris positions can be roughly estimated using AAM algorithm which provides the average positions of the feature points. However, since iris positions are often erroneously located due to the variation in the image, especially due to eyeball movement, an accurate iris detection algorithm is required to estimate the pixel distance.

An iris detecting operator is applied to detect the irises [18]. For fast and accurate iris detection, a region of interest (ROI) is defined as twice the size of rectangle (white solid line in Fig. 4) by the boundary rectangle using four facial features around the iris (white dotted line in Fig. 4), as shown in Fig. 4. In general, since the intensity of the iris is lower than its neighboring regions. Simple mask is designed and applied to utilize this property as shown in Fig. 5. That is, the sum of the intensity differences between the center area (A_0) and the neighboring regions from A_1 to A_8 is computed as follows:

$$D = \sum_{i=1}^{8} (A_i - A_0). \tag{1}$$

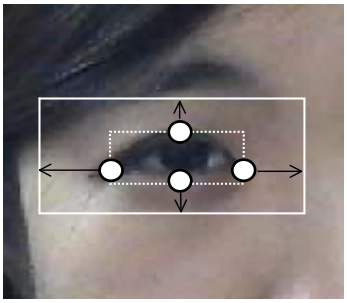
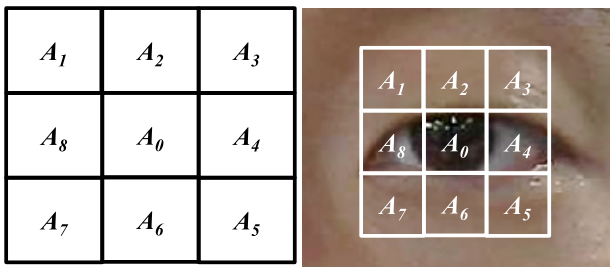


Fig. 4 Selection of the ROI for iris detection.



(a) (b)

Fig. 5 An example of iris detection. (a) A mask for iris-shaped feature extraction. (b) An illustration of the applied mask.

The iris is finally detected by finding the highest score position for D, which corresponds to the darkest area in the ROI.

2.3 Face Pose Estimation

The face rotation affects the pixel distances between irises in different images even when the person is located at the same position as shown in Fig. 6. In order to compensate the variations in face pose, estimation is applied as it is directly related to the accuracy of position estimation. There are multiple approaches to estimate the pose of an object (especially a face) using numerical techniques [26]–[28].

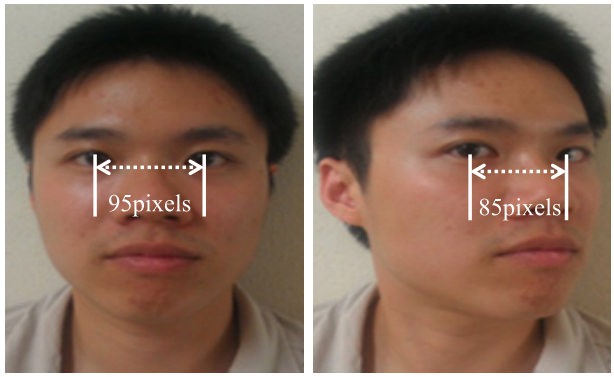
A popular method is the Pose from Orthography and Scaling with Iterations (POSIT) algorithm [28], which converges quickly and does not require an initial estimate of the pose. To estimate the face pose using the POSIT algorithm, same facial feature points as for AAM are used. The positions of irises are then converted onto a face, as viewed from the front, using estimated face pose.

The face of the user can be rotated about three orthogonal axes, as shown in Fig. 7. The rotation matrix of the face pose is represented by the combination of three rotation matrices, represented as follows:

$$R(\alpha,\beta,\gamma) = R_z(\alpha)R_y(\beta)R_x(\gamma), \tag{2}$$

where,

$$R_z(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{3}$$



(a) (b)

Fig. 6 Examples of the distance between eyes difference that depend on face rotation at the same head position. (a) Frontal face. (b) Rotated face (20 degree).

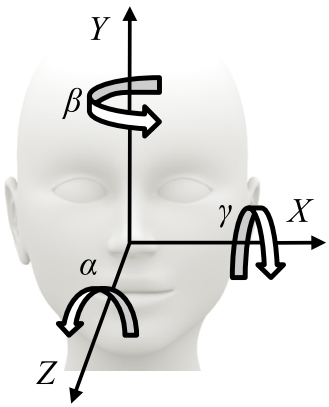


Fig. 7 A principle axis for face rotation.

$$R_y(\beta) = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix}, \tag{4}$$

$$R_x(\gamma) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & -\sin\gamma \\ 0 & \sin\gamma & \cos\gamma \end{pmatrix}. \tag{5}$$

Input positions of the rotated face, $(x_{input}, y_{input}, z_{input})$ are given by POSIT algorithm. Converted positions on the frontal face, $(x_{front}, y_{front}, z_{front})$, are then computed by multiplying the input position and inverse of the rotation matrix, given by:

$$\begin{bmatrix} x_{front} \\ y_{front} \\ z_{front} \end{bmatrix} = R(\alpha,\beta,\gamma)^{-1} \begin{bmatrix} x_{input} \\ y_{input} \\ z_{input} \end{bmatrix}. \tag{6}$$

User position is estimated using inter-pupillary distance between converted irises.

3. Position Estimation in the Real-World

To estimate the user's position in practice, inter-pupillary distance is used. The distance in terms of the number of

pixels (inter-pupillary distance) varies depending on the distance from the camera. Using this relationship, user position can be estimated in real-world. However, this relationship usually depends on the characteristics of the camera. Since it is difficult to obtain/know the camera characteristics, a simple perspective mapping function is used to estimate the user position directly without the need of camera information.

In relevant literature a common non-glass type 3D displays have been designed and implemented with a 6.5 cm inter-pupillary distance [10], [29], the paper use the same length to generate the mapping function for user position estimation, as shown in Fig. 8. This function is generated only once for the camera regardless of the user, and is a set of different widths of inter-pupillary distances (in pixels) at different locations (in centimeters).

Figure 9 shows a power function adopted to model the relationship for a simple perspective mapping function using the inter-pupillary pixel distance in image l_d. The power function is expressed as:

$$Z = \frac{a}{l_d} + b. \tag{7}$$

In (7), the parameters are constant coefficients of the weighting function. These coefficients are estimated by least-square method. The distance along the X-axis is computed by multiplying the pixel movements for the center of the image x, as expressed in (8). (The inter-pupillary distance as 6.5 cm is assumed.)

$$X = x\frac{6.5\,\text{cm}}{l_d}. \tag{8}$$

Although, inter-pupillary distance shows less variation as compared to other features in the face, accuracy of position estimation can further be improved using the actual inter-pupillary distance L_{person} instead of the default value of 6.5 cm, provided as follows:

$$l'_d = \frac{6.5\,\text{cm}}{L_{person}} l_d. \tag{9}$$

Fig. 8 Example of calibration procedure. (a) Inter-pupillary distance is 145 pixels when the user is at 40 cm to the camera. (b) Inter-pupillary distance becomes 98 pixels when the user is at 60 cm away from the camera.

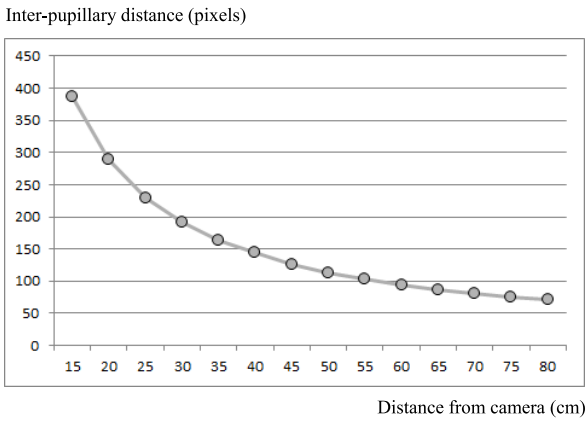


Fig. 9 Relationship between the inter-pupillary width in pixels and the distance from camera in centimeters.

4. Experimental Results

To demonstrate the effectiveness of the proposed method, experiments are performed at specific working distances (from 30 to 70 cm). In the experiments, nine predefined positions are determined to measure the accuracy of position estimation, as shown in Fig. 10.

In the experimental processes, twelve people participated as test subjects. Video sequences were recorded at nine predefined positions for each subject using a common webcam. In total, 108 sets of video sequences were captured for these experiments. When the video sequences were being recorded, the users were able to move their faces in a natural way while looking at the display. The resolution of each video sequence was 640 × 480 at 15 fps. These experiments were performed on a PC with a dual core 2.5-Ghz CPU, and the average computational time was about 15 ms to 30 ms. Actual computation time depended on user movement but was around 20 ms on average.

Figure 11 shows parts of the experimental results. White dots and a red solid line in the figure indicate the facial feature points and the distance between the detected

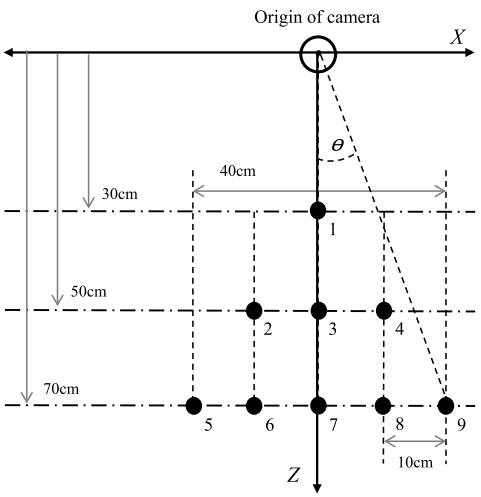


Fig. 10 Predefined positions for the experiments.

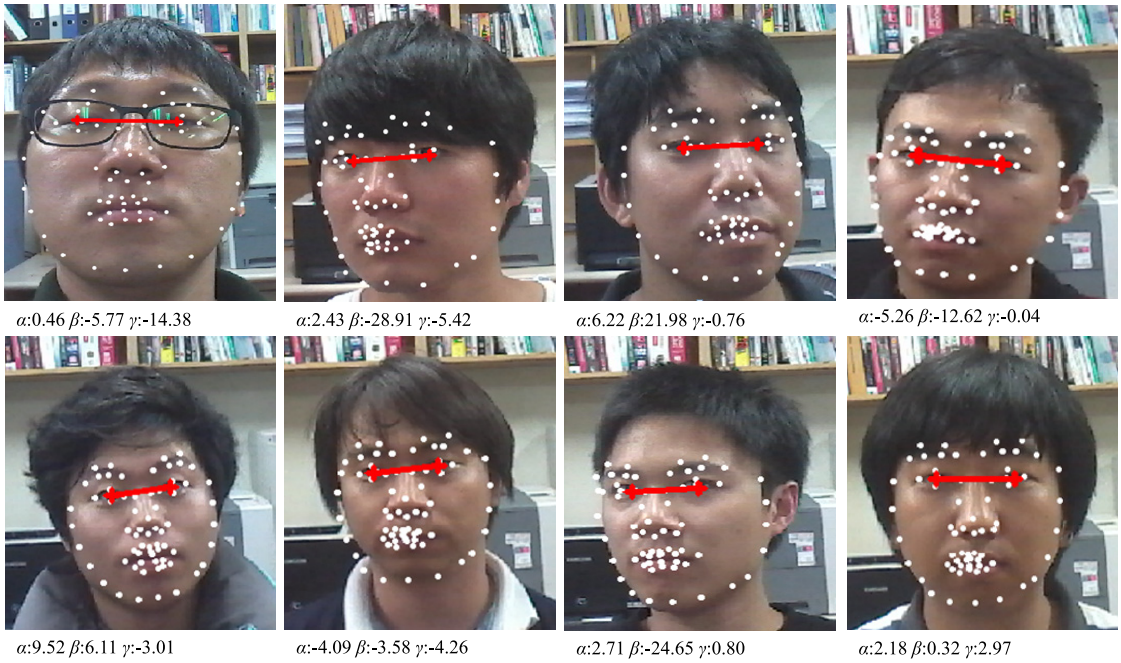


Fig. 11 Examples of the result.

Table 1 Comparison of RMSE.

Approach	*Average RMSE*	
	X	*Z*
Proposed method	0.67 cm	0.81 cm
Without personal calibration	0.78 cm	0.94 cm
Without iris detection	0.82 cm	1.10 cm
Without face pose estimation	1.25 cm	2.12 cm

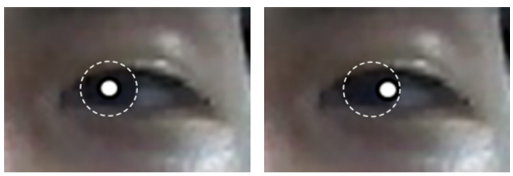


Fig. 12 Comparison of iris location indicated by white dots. (a) By iris detector. (b) By AAM.

Fig. 13 Erroneous examples. (a) Quick head movements of user. (b) Misdetection of iris due to the eyebrows. (c) Error of AAM algorithm.

irises, respectively. Also, pose estimation results for α, β, and γ are shown at the bottom of the each sub-image in the figure.

To measure the accuracy of the position estimation, the root mean square error (RMSE) is calculated and is given as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_{g,i} - p_{e,i})^2}, \tag{10}$$

where n, $p_{g,i}$ and $p_{e,i}$, are the number of positions, ground truth positions and estimated user positions, respectively.

Table 1 shows the RMSE values of distances along X- and Z-axes of the proposed method with personal calibration using the actual distance between irises measured with a ruler. The accuracy without personal calibration (using the average inter-pupillary distance) is slightly lower. In the proposed scheme, average inter-pupillary distance and its standard deviation of test subjects are 6.63 cm and 0.23 cm, respectively. It is also possible to directly estimate iris position by AAM. However, the error tends to increase due to the eyeball movement as shown in Fig. 12. The results show that the estimation error of position is largest when face post estimation is not applied.

Erroneous examples are also shown in Fig. 13. Some error is caused by blurring due to the quick head movements of user, as shown in Fig. 13 (a). Others are caused due to errors of AAM or of the iris detection algorithm, as shown in Figs. 13 (b) and (c).

A comparison of errors among other existing systems based on single camera is shown in Table 2. Kim's method [18] and *FaceAPI* [30] are compared together because of their similarity to the proposed method. Since each method has its own limitations or constraints in its

Table 2 Comparison with other systems using single camera.

Method	*Average RMSE*	
	X	*Z*
Kim's method [18]	1.51 cm	2.50 cm
FaceAPI [30]	1.00 cm	1.00 cm
Proposed method	0.67 cm	0.81 cm

application, their best performances are compared with the proposed scheme. The same input dataset is used as in Kim's method. Although FaceAPI system is commercially available, a functionally limited non-commercial version is used and tested for comparison with the proposed technique. The results demonstrate that the proposed method yielded the lowest error and provided enough accuracy to view 3D contents.

5. Conclusions

In recently years, non-glass type 3D display has become more popular. For these displays, an accurate user position estimation algorithm becomes essential as their quality is determined by the accuracy of the estimated user position.

In this paper, an accurate user position estimation method using a single camera for non-glass type 3D displays has been proposed. For this purpose, a simple and efficient method is developed to estimate the user position. To compensate the variations in pose, face pose estimation algorithm is applied. In order to further improve the accuracy, the irises of the user are detected. For position estimation in the real-world, a perspective mapping function is applied using the average inter-pupillary distance. Also, personal inter-pupillary distances can be used for better accuracy.

Experiments are carried out to verify the performance of the proposed user position estimation method. The results show that the proposed method has better performance as compared to its predecessors for non-glass type 3D displays.

Acknowledgments

This work is supported by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2012-H0301-12-1011).

References

[1] S. Pastoor and M. Wopking, "3-D displays: A review of current technologies," Displays, vol.17, pp.100–110, 1997.
[2] P. May, "A survey of 3-D display technologies," Information Display, vol.32, pp.28–33, 2005.
[3] D. Hutchinson and H. Neal, "The design and implementation of a stereoscopic microdisplay television," Consum. Electron., vol.54, pp.254–261, 2008.
[4] N. Holliman, 3D display systems, Handbook of Opto-electronics, IOP Press, 2005.
[5] F. de Sorbier, Y. Uematsu, and H. Saito, "Depth camera based system for auto-stereoscopic displays," 20th International Conference on Artificial Reality and Telexistence (ICAT 2010), pp.184–188, 2010.
[6] D. Sandin, T. Margolis, J. Ge, J. Girado, T. Peterka, and T. DeFanti, "The varrier autostereoscopic virtual reality display," Proc. ACM SIGGRAPH '05/ACM Trans. Graphics, vol.24, no.3, pp.894–903, 2005.
[7] L. Lipton and M. Feldman, "A new autostereoscopic display technology: The SynthaGram," Proc. SPIE Photonics West: Electronic Imaging, 2002.
[8] N.A. Dodgson, "Autostereoscopic 3D displays," Computer, vol.38, no.8, pp.31–36, 2005.
[9] D. Ezra, G.J. Woodgate, B.A. Omar, N.S. Holliman, J. Harrold, and L.S. Shapiro, "New autostereoscopic display system," Proc. SPIE, vol.2409, pp.31–40, 1995.
[10] C.-H. Tsai, P. Lai, K. Lee, and C.K. Lee, "Fabrication of a large F-number lenticular plate and its use as a small-angle flat-top diffuser in autostereoscopic display screens," Proc. SPIE, vol.3957, pp.322–329, 2000.
[11] A. Boev, A. Gotchev, and K. Egiazarian, "Crosstalk measurement methodology for auto-stereoscopic screens," Proc. 3DTV Conference, pp.1–4, 2007.
[12] Q. Mühlbauer, K. Kühnlenz, and M. Buss, "A model-based algorithm to estimate body poses using stereo vision," Proc. 17th International Symposium on Robot and Human Interactive Communication, pp.285–290, 2008.
[13] Y. Yoon, G. DeSouza, and A. Kak, "Real time tracking and pose estimation for industrial objects using geometric features," IEEE Int. Conf. On Robotics & Automation, pp.3473–3478, 2003.
[14] MS Kinect, http://www.microsoft.com/en-us/kinectforwindows/, 2012.
[15] S. Kawahito, I.A. Halin, T. Ushinaga, T. Sawada, M. Homma, and Y. Maeda, "A CMOS time-of-flight range image sensor with gates-on-field-oxide structure," IEEE Sensors J., vol.7, no.12, pp.1578–1586, 2007.
[16] P. Harmann, "Retroreflective screens and their application to autostereoscopic displays," Proc. SPIE, vol.3012, pp.145–153, 1997.
[17] E. Murphy-Chutorian and M. Trivedi, "Head pose estimation in computer vision: A survey," IEEE Trans. Pattern Anal. Mach. Intell., vol.31, no.4, pp.607–626, 2009.
[18] B. Kim, H. Lee, and W.-Y. Kim, "Rapid eye detection method for non-glasses type 3D display on portable devices," Consum. Electron., vol.56, no.4, pp.2498–2505, 2010.
[19] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models," IEEE Trans. Pattern Anal. Mach. Intell., vol.23, no.6, pp.681–685, 2001.
[20] L. Matthews and S. Baker, "Active appearance models revisited," Int. J. Comput. Vis., vol.60, no.2, pp.135–164, 2004.
[21] Y. Ma and X. Ding, "Robust real-time face detection based on cost-sensitive AdaBoost method," ICME '03, vol.2, pp.465–468, 2003.
[22] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," IEEE Conference on Computer Vision and Pattern Recognition, vol.1, pp.511–518, 2001.
[23] Y. Cui and Z. Jin, "Facial feature points tracking based on AAM with optical flow constrained initialization," J. Pattern Recognition Research, vol.7, pp.72–79, 2012.
[24] T.F. Cootes, G.V. Wheeler, K.N. Walker, and C.J. Taylor, "View-based active appearance models," Image Vis. Comput., vol.20, pp.657–664, 2002.
[25] B.D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," Proc. DARPA Image Understanding Workshop, pp.121–130, 1981.
[26] D.G. Lowe, "Fitting parameterized three-dimensional models to images," IEEE Trans. Pattern Anal. Mach. Intell., vol.13, no.5, pp.441–450, 1991.
[27] J.S.-C. Yuan, "A general photogrammetric method for determining

object position and orientation," IEEE Trans. Robot. Autom., vol.5, no.2, pp.129–142, 1989.
[28] D.F. Dementhon and L.S. Davis, "Model-based object pose in 25 lines of code," Int. J. Comput. Vis., vol.15, no.1, pp.123–141, 1995.
[29] C.Y. Chen, Q.L. Deng, and H.C. Wu, "A high-brightness diffractive stereoscopic display technology," Displays, vol.31, pp.169–174, 2010.
[30] FaceAPI, http://www.seeingmachines.com/product/faceapi/, Aug. 2012.

Byeoung-su Kim received his B.S. degree from Hanyang University, Seoul, Korea in 2006. He is currently working toward a Ph.D. in the Department of Electronics and Computer Engineering at the same university. His research interests include intelligent video surveillance, 3D display techniques, face recognition, and pattern recognition.

Cho-il Lee received his B.S. degree from Sahmyook University, Seoul, Korea in 2009. He is currently working toward an M.S. degree in the Department of Electronics and Computer Engineering at Hanyang University. His research interests include intelligent vehicle design and 3D display techniques.

Seong-hwan Ju received the B.S. degrees in Computer Engineering from Halla University, Wonju, Korea in 2004. He received his M.S. degree in Mechatronics Engineering from Hanyang University, Seoul, Korea in 2007. He is currently working at 3D technology department, LG Display.

Whoi-Yul Kim received his B.S. degree in Electronic Engineering from Hanyang University, Seoul, Korea in 1980. He received his M.S. from Pennsylvania State University, University Park, Pennsylvania in 1983 and his Ph.D. from Purdue University, West Lafayette, Indiana in 1989, both in Electrical Engineering. From 1989 to 1994, he was with the Erick Johnson School of Engineering and Computer Science at the University of Texas at Dallas. Since 1994, he has been on the faculty of Electronic Engineering at Hanyang University, Seoul, Korea. He has been involved with research and development of 3D vision systems.